Argument Identification in a Language Without Labeled Data

João Rodrigues(✉) and António Branco

NLX—Natural Language and Speech Group, Department of Informatics Faculdade de Ciências, University of Lisbon, 1749-016 Campo Grande, Lisboa, Portugal
{joao.rodrigues, antonio.branco}@di.fc.ul.pt

Abstract. This paper addresses the issue of how to obtain processing tools for argument identification for the vast majority of the languages that, differently from English, have little to no relevant labeled data.

This issue is addressed by taking an under-resourced language as a case study, namely Portuguese, and by experimenting with three techniques to cope with the scarceness of data: to obtain labelled data by machine translating data sets from another language labelled with respect to argument identification; to transfer to the argument identifier the language knowledge captured in distributional semantic models obtained during the resolution of other tasks for which more data exist; to expand data for argument identification with text augmenting techniques.

The results obtained demonstrate that it is possible to develop argument identification tools for under-resourced languages with a level of performance that is competitive to the ones for languages with relevant language resources.

Keywords: Argument identification · Argument mining · Machine translation · Learning transfer · Data augmentation · Under-resourced languages · Portuguese

1 Introduction

Automatic argument mining may support a number of high-level applications, including argument search, decision making, automated reasoning, or user review analysis, among several others. As a consequence, there has been an increasing interest on the research about argument mining, which is visible in the range of shared tasks that have been addressed in the last editions of the SemEval workshop [10].

The language processing task of argument mining faces a number of challenges, among which a most notorious one is the lack of a widespread consensus about the most appropriate analysis of arguments. Arguments have different components, e.g. premises and claims, may be related under different possible relations, e.g. attack, support, etc., and have different levels of quality, e.g. persuasiveness, convincing, etc. Many different analysis for arguments have been

P. Quaresma et al. (Eds.): PROPOR 2020, LNAI 12037, pp. 335–345, 2020.
https://doi.org/10.1007/978-3-030-41505-1_32

proposed in the literature, with different types of components and relations [31], and with different argumentation schemes and models, i.e. patterns of propositions that form an argument—for which the Walton [33] schemes and Toulmin [32] models are two well known proposals.

There has been nevertheless a reasonable consensus in the literature that the overall task of argument mining can be usefully broken down into a chain of subsidiary sub-tasks that include, for instance, argument identification, component identification, relation extraction and argument quality assessment [31].

As the mainstream techniques to handling argument mining are mostly based on machine learning, another notorious challenge for argument mining concerns the scarcity of data, and in particular of data sets conveniently annotated with the information about the components, quality, etc., of arguments. Language resources with good quality for argument mining are expensive to develop—requiring the manual labor of annotating massive amounts of data —, and while a few currently exist for English, very little is available yet for the vast majority of the other approximately 7,000 languages in the world.

In this paper, we focus on the initial sub-task of automatic argument mining, namely on argument identification, which consists of taking as input a segment of text and returning whether it is an argument or not. We report on the results of applying a number of approaches that may help to address the scarceness of data for argument mining in a language that is less-resourced than English. As a case study of an under-resourced language in this respect, we consider Portuguese [2–4], for which very little data is available yet for argument mining [23].

Concerning the task of argument identification, we pursue here a twofold goal. On the one hand, by resorting to machine translation [16,24,27], we report on performing argument identification with state of the art techniques over a data set in Portuguese that results from translating a mainstream data set in English annotated with argument identification.

On the other hand, taking that translated data set as a basis, we report on a number of subsequent experiments with approaches that seek to further mitigate the data scarceness in argument identification. We will report on transfer learning from distributional semantic models, also known as word embeddings, obtained from data sets of Portuguese that are much larger than the data set obtained for argument identification via translation.

We will report also on further experiments to mitigate data scarcity by resorting to a range of data augmenting techniques, from the simpler one of randomly inserting QWERTY characters, to the more complex one of generating segments with the help of Transformer-based models (BERT [7] and GPT-2 [22]).

Experiments and results presented in this paper demonstrate that, having Portuguese as case study, the approaches we propose provide substantive enhancements for argument mining in languages that are under-resourced in terms of data sets relevant for argument identification, allowing to develop argument identifiers whose performance is competitive to the ones for languages with relevant language resources, like English.

2 Related Work

Like the study in [23], our work also addresses an argument mining task in Portuguese and apply techniques to mitigate the lack of annotated resources, including by applying triangulation supported by machine translation [6]. However, while we address the sub-task of argument identification, [23] addressed the sub-task of argumentative relation identification. These authors used a cross-lingual setting between a source data set in English and a target data set in Portuguese, and explored two techniques, a projection and a direct-transfer technique. In the projection technique, the (English) source data set was automatically translated and a machine learning algorithm was trained on the (Portuguese) target data set. In the direct-transfer technique, a machine learning algorithm was trained on the source data and fine-tuned to the target data.

Several publications have reported on experiments that resorted to distributional semantic models aiming at enhancing several language processing tasks, [19,25,26] including argument mining for English. These kind of models have been used as features to predict argument structure [18,21] and argument convincingness [29], or used in the embedding layer of neural networks to identify attack or support argumentative relations [5,15]. The enhancement of machine learning algorithms with different distributional semantic models has been evaluated also on an argument component identification task [9] by comparing two neural network architectures (Convolutional and Long-Short Term Memory) that were tested with three distributional representations, the word2vec [20], the dependency-based embeddings [17] and factuality/certainty-indicating embeddings. To the best of our knowledge, the present paper is the first to report on the argument identification task for Portuguese enhanced with transfer learning from distributional semantic models and with the use of transformer-based language models as generators of augmented data.

3 Experiments

To address our research goals, three main experiments are undertaken.[1] The first one relies on an intermediate translation step. An English annotated data set, created specifically for the argument identification task, is automatically translated into Portuguese. Over these two data sets, a state-of-the-art machine learning algorithm, namely BiLSTM, is applied to the argument identification task. Having the performance scores for the two data sets will permit us to have an insight into how much the noise introduced by the machine translation procedure affects the argument identification task in the target language. In other words, this will permit us to have an insight into how useful is this approach to address argument identification in a language that is under-resourced for this task.

[1] The source code to reproduce these experiments are available at https://github.com/nlx-group/argument-identification.

The second experiment explores transfer learning from distributional semantic models for Portuguese to enhance the machine learning model in the argument identification task obtained in the first experiment.

Finally, the third experiment consists of applying a range of different data augmentation techniques, including the generation of data improved by the fine-tuning of a transformer-based language model.

3.1 First Experiment: Machine Translation

The corpus used in the three experiments for Portuguese is obtained from a mainstream data set in English, the UKP Sentential Argument Mining Corpus [30], which is translated into Portuguese by resorting to Google Translate. Examples from this data set are presented in Table 1.

The English data set was created by including texts on eight controversial topics in each one of these domains: news, editorials, blogs, debate forums, and encyclopedia articles. By taking into account the respective topic, each sentence in the corpus was manually annotated as being an argument or a non-argument. This corpus has approximately 25k sentences of which 10k are labeled as arguments and 15k as non-arguments. The definition for argument followed by the annotators was *a span of text expressing evidence or reasoning that can be used to either support or oppose a given topic.*

Table 1. Sample from the data set: the first two sentences are from the UKP corpus; the last two are their Portuguese Google translations.

Sentence	Label
We need a safe, genuinely sustainable, global and green solution to our energy needs, not a dangerous diversion like nuclear power	*argument*
There are many notable authors of books and articles that render scientific findings available in lay language to a wider public	*non-arg*
Precisamos de uma solução segura, genuinamente sustentável, global e verde para nossas necessidades de energia, não de uma diversão perigosa como a energia nuclear	*argument*
Existem muitos autores notáveis de livros e artigos que disponibilizam descobertas científicas em linguagem leiga para um público mais amplo	*non-arg*

For the Portuguese data set, we adopted the same split proportion as in the original (English) data set, that is, 70% of the total instances for training, 10% for development and 20% for testing.

As the machine learning approach to address the argument identification task, we implemented in Tensorflow [1] a Bidirectional Long Short-Term Memory

(BiLSTM) neural network [12,28]. This network used a trainable embedding layer for the input, instantiated with random embeddings using the FastText [14] 1M words as the vocabulary, that was followed by a single BiLSTM layer with 48 units, and was used to tune a model with a hyper-parameters grid search.[2]

When experimented with the English data set, this set up obtained a performance for argument identification in line with the state of the art, namely an F-measure of 0.7220, which compares very competitively with the F-measure of 0.6580 obtained from the state-of-the-art models reported in [30].

A reason to explain this difference is the following. Although both models use the same data set and split proportions for training, development and testing, in [30] the aim was to evaluate cross-topic argumentation mining, cross-testing each of the 8 topics by training in 7 of them and evaluating on the eighth topic. In our work, in turn, we did not aim at a cross-topic approach. We randomize the data set before splitting it, and by training with data also on the same topic a higher F-measure was obtained.

We applied the same machine learning algorithm for the Portuguese data set. The same procedures were repeated, using the same hyper-parameters and a randomized embedding layer with the Portuguese version of FastText 1M words. We obtained an F-measure of 0.7228, which indicates a competitive performance when compared to the English counter-part (0.7220) developed under the same settings but with data sets specifically annotated for this language. We evaluated the same model on 60 manually reviewed sentences from the test set and obtained a delta of 0.0320 in comparison with the machine translated output. This leads us to believe that the resulting data set for Portuguese is suitable for the argument identification task in Portuguese, and thus that (machine) translation may be a good enough option for under-resourced languages in what concerns obtaining labeled data for the development of argument identifiers.

3.2 Second Experiment: Learning Transfer

Aiming at enhancing the performance of the argument identification tool, we sought to transfer knowledge to the respective machine learning classifier from different word embeddings of Portuguese.

We choose to experiment with semantic spaces of different natures, namely from: (a) FastText [14], a distributional semantic space that takes morphological information into account; (b) multilingual word vectors [8], which are jointly learned embeddings from parallel data; and (c) GloVe and word2vec models created with the models from STIL2017 [11]. All the distributional semantic models had the same vector size of 300 units.

[2] The best hyper-parameters were: 10 epochs, batch size of 64, sequence length of 30, learning rate of 0.01, dropout of 0.8, 48 LSTM units and a softmax cross-entropy with an Adam optimization. We manually experimented with the grid search values, and used an early stop technique using the best F-measure obtained from the development data set. An 0.7576 F-score was obtained on this data set from the average of 10 runs. Each trial took approximately 10 min in a GeForce RTX 2080 GPU.

Table 2. Second Experiment - Performance of the argument identifier enhanced with learning transfer techniques, measured with Accuracy, F-measure, Precision and Recall.

Model	Accuracy	F-measure	Precision	Recall
Baseline	0.6734	0.7228	0.6872	0.7627
Word embeddings				
Fasttext	**0.7056**	0.7433	0.7252	0.7640
CMU	0.6984	0.7338	0.7239	0.7449
GloVe	0.7038	0.7399	0.7263	0.7553
CBow	0.6984	0.7370	0.7200	0.7579
Contextual word embeddings				
BERT	**0.7588**	0.7580	0.8619	0.6764

We created a different model with each one of these three semantic spaces encoded in the word embedding layer of the neural network. In all three models, this layer was non-trainable, that is the weights for the embeddings were fixed during all the learning phases. Thus, all of the learning parameters resided solely on the parameters found in the BiLSTM layer, obtained from the baseline hyper-parameter grid search.

Given the latest transformer-based architectures, such as BERT [7], have been the state of the art in several natural language processing downstream tasks, we also experimented with the transfer of knowledge from a BERT fine-tuned for Portuguese. BERT is a bidirectional encoder that learns from the left and right context for each word in a sentence and is trained on two tasks: a mask language task and a sentence prediction task. While in the previous distributional semantic models, the neural network has an embedding layer encoding the respective semantic space, BERT is itself a neural network with the semantic space encoded through several neural network layers.

We fine-tuned a pre-trained multi-language BERT model resorting to adapters.[3] An adapter [13] re-purposes a neural network model by adding a new neural network layer, typically a top-layer, and while the original neural layers are kept frozen, the new layer is fine-tuned. This approach reduces the number of parameters necessary for retraining a model thus achieving a faster convergence. Here the training and development sets were used for the fine-tuning.

The results obtained are displayed in Table 2. The baseline is the argument identifier obtained in the first experiment, with an accuracy of 0.6734. All the other models surpassed this baseline. The fine-tuned BERT model outperformed

[3] We used the *multi_cased_L-12_H-768_A-12* model. The best hyper-parameters found were a maximum sequence length of 128, a learning rate of 2e-4, 8 training epochs, and a batch size of 32. We manually experimented with the grid search values of the maximum sequence length, learning rate and the number of training epochs. Each trial took approximately 25 min using a GeForce RTX 2080 GPU.

all the other models, with an accuracy of 0.7588, more than 8 points higher than the baseline score.[4]

3.3 Third Experiment: Data Augmentation

Seeking to further enhance the performance of the argument identifier, we experimented with data augmentation techniques. To obtain a series of comparable results, we adopted the base BiLSTM classifier (with the hyper-parameters initially tuned) for all experiments, except when BERT was used, which has its own neural architecture and parameters. To keep with the series of experiments, we resorted to the model with word embeddings that led to the best improvement of the BiLSTM performance in the second set of experiments. We resorted to the Fasttext as the base setup for the third set of experiments, and to its accuracy as the baseline performance.

The performance of several models was investigated, that were obtained in several data augmentation exercises. In each of these exercises, the generated data was added individually to the original training data set. We used the same hyper-parameters as in the previous experiments. The labels (argument or non-argument) of the synthetic sentences are made identical to the labels of the respective base sentences.

First, we resorted to data augmentation techniques that involve the handling of characters: (a) for each randomly picked character c in text, concatenate to it a $QWERTY$ character, which corresponds to a key of the $QWERTY$ layout keyboard that is a neighbor of the key corresponding to character c; (b) for each randomly picked position in text, concatenate a *randomly* picked character; and (c) *delete* characters at random in text.

Second, we used techniques that involve the handling of words. For each word w randomly picked in text: (a) *insert* another word after it; and (b) *replace* it by another word. The new word, to be inserted, is a most semantically similar word to w, where semantic similarity is determined by the smallest cosine of the angles between vectors of words in a distributional semantic space. Three semantic spaces were experimented with, namely Fasttext, GloVe and BERT.

Finally, we resorted to synthetic data where sentences are generated with the help of a language model, namely the GPT-2 model [22]. Each sentence in the original training data is used as the context for GPT-2 to generate three other (synthetic) sentences. GPT-2 is a large transformer-based language model trained on the word prediction task from a 40 GB of web data corpus. It outperforms several other language models on a number of language tasks, thus being a good option to generate text.

Given that the original models of GPT-2 were trained with English corpora, we trained three *355M parameter* models for Portuguese, with three Portuguese corpora from different domains: Wikipedia;[5] CetemPúblico, with articles from

[4] It is worth recalling that on pair with the baseline, BERT is the only other model taking advantage of a fine-tuning and hyper-parameter grid-search.

[5] Portuguese Wikipedia data dump of 01/09/2015.

Table 3. Third Experiment - Performance of the argument identifier enhanced with data augmentation techniques, measured with Accuracy, F-measure, Precision and Recall.

Augmentation	Accuracy	F-measure	Precision	Recall
Baseline	0.7056	0.7433	0.7252	0.7640
Character handling				
Insert QWERTY	0.7019	0.7175	0.7619	0.6788
Insert random	**0.7087**	0.7315	0.7541	0.7116
Delete random	0.7025	0.7201	0.7588	0.6858
Word handling				
Insert Fasttext	**0.7169**	0.7491	0.7419	0.7570
Insert GloVe	0.7150	0.7500	0.7366	0.7665
Insert BERT	0.7146	0.7494	0.7361	0.7649
Replace Fasttext	0.7078	0.7328	0.7498	0.7185
Replace GloVe	0.7145	0.7436	0.7473	0.7427
Replace BERT	0.7112	0.7285	0.7678	0.6947
Sentence handling				
Generate Wikipedia	**0.7174**	0.7479	0.7453	0.7513
Generate CetemPúblico	0.7071	0.7390	0.7390	0.7423
Generate Europarl	0.7079	0.7300	0.7567	0.7091

the Público newspaper; and Europarl, with transcriptions of debates from the European Parliament. We used the gpt-2-simple module.[6] Accordingly, the initial data set with 25 k sentences (translated into Portuguese) doubled in size to 50 k sentences with each character or word handling technique for data augmentation experimented with, and quadrupled to 100 k sentences with each sentence handling technique.

The results are presented in Table 3. Every technique experimented with led to improved performance of the argument identifier, except in the cases of the *QWERTY* and *Delete* exercises, yet only with a slight decay with respect to the baseline with an accuracy score of 0.7056. The best solution is obtained with GPT-2 trained with Wikipedia, scoring 0.7174 accuracy.

The gain of over 1 accuracy point seems to indicate that the advantage obtained by having more data only modestly offsets the noise introduced by labeling the generated sentences, as an argument or a non-argument, with the same label of their context sentences.

[6] gpt-2-simple was obtained from https://github.com/minimaxir/gpt-2-simple. We used a generation length of 60 units, the top 3 tokens and one sample per sentence. The training of each model and respective data generation took approximately 3 days using a GeForce RTX 2080 GPU.

4 Conclusions

In this paper, we address the issue of how to obtain argument identification tools for the vast majority of the approximately 7,000 languages of the world that, differently from English, have little to no labeled data that permits the training of solutions for this language processing task. We sought to tackle this issue by taking a language that is under-resourced for this task as a case study, namely Portuguese, and by experimenting with three types of techniques to cope with the scarceness of curated data: to obtain (seed) data by machine translating data from other languages labelled with respect to argument identification (i.e. from English); to transfer to the (seed) argument identifier the language knowledge captured in distributional semantic models (word embeddings) obtained with other language processing tasks for which more data exist; to augment the seed data (initially obtained by translation) with techniques that transform it into new versions of them by handling characters, words or sentences.

The results obtained demonstrate that it is possible to obtain argument identification tools for under-resourced languages with a level of performance (0.7228 F-score) that is competitive with the performance of the tools for the languages with relevant resources (0.7220 F-score for English under the same experimental settings), by translating the later and then training a (BiLSTM based) argument identifier on the output, in the target under-resourced language.

They demonstrate also that some performance gains can be obtained, though somewhat modest (over 1 accuracy point), with data augmenting techniques, with sentence handling techniques contributing better than word handling ones, which in turn contribute better than character handling ones.

The results of the experiments undertaken demonstrate also that it is possible to improve the performance of the seed identifier by transferring the language learning captured in distributional semantic models obtained during the training for other language processing tasks that may resort only to unlabelled data. As expected, contextual word embeddings support larger improvements (over 12 accuracy points) than non-contextual word embeddings (over 3 accuracy points) over the 0.6734 accuracy baseline. Concomitantly, these experiments happen also to set the state of the art in 0.7588 accuracy for argument identification in Portuguese.

Acknowledgments. The research reported here was partially supported by PORTULAN CLARIN—Research Infrastructure for the Science and Technology of Language, funded by Lisboa 2020, Alentejo 2020 and FCT—Fundação para a Ciência e Tecnologia under the grant PINFRA/22117/2016.

References

1. Abadi, M., et al.: Tensorflow: large-scale machine learning on heterogeneous systems (2015). https://www.tensorflow.org/, software available from tensorflow.org
2. Branco, A.: A língua portuguesa face ao choque tecnológico digital. Revista do Instituto Internacional da Língua Portuguesa **2**, 28–36 (2013)

3. Branco, A.: We are depleting our research subject as we are investigating it: in language technology, more replication and diversity are needed. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
4. Branco, A., et al.: The Portuguese Language in the Digital Age. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29593-5
5. Cocarascu, O., Toni, F.: Identifying attack and support argumentative relations using deep learning. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1374–1379 (2017)
6. Costa, F., Branco, A.: TimeBankPT: a timeML annotated corpus of Portuguese. In: Proceedings of the International Conference on Language Resources and Evaluation, pp. 3727–3734 (2012)
7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186 (2019)
8. Ferreira, D.C., Martins, A.F., Almeida, M.S.: Jointly learning to embed and predict with multiple languages. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2019–2028 (2016)
9. Guggilla, C., Miller, T., Gurevych, I.: CNN-and LSTM-based claim classification in online user comments. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2740–2751 (2016)
10. Habernal, I., Wachsmuth, H., Gurevych, I., Stein, B.: SemEval-2018 task 12: the argument reasoning comprehension task. In: Proceedings of the 12th International Workshop on Semantic Evaluation, pp. 763–772 (2018)
11. Hartmann, N., Fonseca, E., Shulby, C., Treviso, M., Silva, J., Aluísio, S.: Portuguese word embeddings: evaluating on word analogies and natural language tasks. In: Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology, pp. 122–131 (2017)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
13. Houlsby, N., et al.: Parameter-efficient transfer learning for NLP. In: International Conference on Machine Learning, pp. 2790–2799 (2019)
14. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pp. 427–431 (2017)
15. Koreeda, Y., Yanase, T., Yanai, K., Sato, M., Niwa, Y.: Neural attention model for classification of sentences that support promoting/suppressing relationship. In: Proceedings of the Third Workshop on Argument Mining (ArgMining2016), pp. 76–81 (2016)
16. Kuang, S., Li, J., Branco, A., Luo, W., Xiong, D.: Attention focusing for neural machine translation by bridging source and target embeddings. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1767–1776 (2018)
17. Levy, O., Goldberg, Y.: Dependency-based word embeddings. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, vol. 2, pp. 302–308 (2014)

18. Li, L., Mao, L., Chen, M.: Word embedding and topic modeling enhanced multiple features for content linking and argument/sentiment labeling in online forums. In: Proceedings of the MultiLing 2017 Workshop on Summarization and Summary Evaluation Across Source Types and Genres, pp. 32–36 (2017)
19. Maraev, V., Saedi, C., Rodrigues, J., Branco, A., Silva, J.: Character-level convolutional neural network for paraphrase detection and other experiments. In: Filchenkov, A., Pivovarova, L., Žižka, J. (eds.) AINL 2017. CCIS, vol. 789, pp. 293–304. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-71746-3_23
20. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
21. Pathak, A., Goyal, P., Bhowmick, P.: A two-phase approach towards identifying argument structure in natural language. In: Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications, pp. 11–19 (2016)
22. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8), 9 (2019)
23. Rocha, G., Stab, C., Cardoso, H.L., Gurevych, I.: Cross-lingual argumentative relation identification: from English to Portuguese. In: Proceedings of the 5th Workshop on Argument Mining, pp. 144–154 (2018)
24. Rodrigues, J.A., Rendeiro, N., Querido, A., Štajner, S., Branco, A.: Bootstrapping a hybrid MT system to a new language pair. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), pp. 2762–2765 (2016)
25. Rodrigues, J.A., Saedi, C., Maraev, V., Silva, J., Branco, A.: Ways of asking and replying in duplicate question detection. In: Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (* SEM 2017), pp. 262–270 (2017)
26. Saedi, C., Rodrigues, J., Silva, J., Maraev, V., et al.: Learning profiles in duplicate question detection. In: IEEE 18th International Conference on Information Reuse and Integration (IEEE IRI 2017), pp. 544–550. IEEE (2017)
27. Santos, R., Silva, J., Branco, A., Xiong, D.: The direct path may not be the best: Portuguese-Chinese neural machine translation. In: Moura Oliveira, P., Novais, P., Reis, L.P. (eds.) EPIA 2019. LNCS (LNAI), vol. 11805, pp. 757–768. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30244-3_62
28. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Trans. Signal Process. **45**(11), 2673–2681 (1997)
29. Simpson, E., Gurevych, I.: Finding convincing arguments using scalable Bayesian preference learning. Trans. Assoc. Comput. Linguist. **6**, 357–371 (2018)
30. Stab, C., Miller, T., Schiller, B., Rai, P., Gurevych, I.: Cross-topic argument mining from heterogeneous sources using attention-based neural networks. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3664–3674 (2018)
31. Stede, M., Schneider, J.: Argumentation mining. Synth. Lect. Hum. Lang. Technol. **11**(2), 1–191 (2018)
32. Toulmin, S.E.: The Uses of Argument. Cambridge University Press, Cambridge (2003)
33. Walton, D.: Argumentation theory: a very short introduction. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 1–22. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-98197-0_1